DOCUMENT RESUME

ED 021 627 PS 001 076

By- Zimiles, Herbert; Silk, Stephen

HEAD START EVALUATION AND RESEARCH CENTER. PROGRESS REPORT OF RESEARCH STUDIES 1966 TO 1967. DOCUMENT 5, COMPARATIVE ITEM-CONTENT ANALYSIS OF ACHIEVEMENT TEST PERFORMANCE IN YOUNG CHILDREN.

Bank Street Coll. of Education, New York, N.Y.
Spons Agency- Office of Economic Opportunity, Washington, D.C.
Pub Date Dec 67
Note- 15p.
EDRS Price MF-$0.25 HC-$0.68
Descriptors- ACHIEVEMENT TESTS, COGNITIVE ABILITY, *COMPARATIVE ANALYSIS, COMPENSATORY EDUCATION, CULTURAL DIFFERENCES, *CULTURALLY DISADVANTAGED, DATA ANALYSIS, GRADE 1, *ITEM ANALYSIS, *MIDDLE CLASS, *TEST RESULTS
Identifiers- Head Start, Metropolitan Achievement Tests, New York State Readiness Tests

To contribute to a more differentiated description of the cognitive status of young disadvantaged children, a comparative study was made of achievement test performance of disadvantaged and middle class groups of children. On the basis of published reading test scores, test data were gathered from nine public schools, four of which were selected from among the highest scoring in each of four school districts. The remaining five were selected as representative of conspicuously low scoring schools. The children from the high scoring schools were predominantly from white middle class families. The children from the low scoring schools were predominantly from lower class Negro or Puerto Rican families. Data analysis consisted of an item-by-item comparison of the performances of the two groups on each of the six subtests of the New York State Readiness Test. The smallest differences in the high and low scoring groups occurred on the listening , matching, and copying subtests. The greatest differences occurred on the numbers subtest. The pervasive differences found to exist in every area of intellectual functioning suggest the need for a program designed to promote general intellectual growth at a basic and integrative level. Statistical data are presented in two tables, and a summary of future research to be conducted is included in the text of this progress report. (JS)

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

HEAD START EVALUATION AND RESEARCH CENTER

Progress Report of Research Studies
1966 to 1967
(Documents 1 - 6)

Document 5

COMPARATIVE ITEM-CONTENT ANALYSIS OF ACHIEVEMENT TEST PERFORMANCE IN YOUNG CHILDREN

Staff

Herbert Zimiles, Ph.D. (Principal Investigator)
Stephen Silk, B.S.

Research Division
Bank Street College of Education
216 West 14th Street
New York, N.Y. 10011

December, 1967

PS001076 ED021627

COMPARATIVE ITEM-CONTENT ANALYSIS OF ACHIEVEMENT TEST PERFORMANCE IN YOUNG CHILDREN

The purpose of this study is to contribute a more differentiated description of the cognitive status of young disadvantaged children through a comparative analysis of achievement test performance of disadvantaged and middle-class groups of children. The analysis focuses on item-by-item differences in performance in order to identify areas of comparative strength and weakness on the basis of the study of item content. In so doing, the study attempts to contribute to an evaluation of the usefulness of standard group testing procedures conducted in the schools as a source of information regarding the areas of achievement and deficit among its children. Although achievement tests are routinely administered periodically in most school systems, their results contribute little specific knowledge about the cognitive functioning of the children. In the case of schools in deprived areas, for example, the results of achievement test administration are cited to indicate how many years behind the children are in reading and other academic subjects. Seldom is an attempt made to go beyond grade equivalent scores in order to obtain diagnostic information, to gauge areas of strength and weakness so that a more differentiated statement of the disadvantaged child's skills and deficits is achieved. The present study represents an attempt to secure such information from achievement test data obtained from children in the first three grades. It is reasoned that the preparation for later schooling conducted by Head Start should be based upon a full knowledge of the capabilities of disadvantaged children during their first few years of grade school.

Method

Achievement test data were gathered from nine public schools. Four pairs of schools, one pair from each of four boroughs, were selected on the basis of published reading test scores, such that one member of the pair was among the highest

scoring schools in the borough while the other member was among the lowest scoring schools. Both members of each pair of schools were located in the same school district, but the children from the high scoring schools were predominantly from white middle-class families whereas the low scoring schools were made up almost entirely of children from lower-class Negro or Puerto Rican families. Test scores were also made available from a ninth school, located in Harlem, thereby enlarging the sample of disadvantaged children. Data are available from the following tests:

New York State Readiness Tests
Metropolitan Achievement Tests:
  Primary I Reading Test, Form B
  Upper Primary Reading Test, Form B
  Upper Primary Arithmetic Test, Form B
  Elementary Arithmetic Test, Form B

It should be noted, however, that data from all the tests cited above were not available from all schools. In Table 1 below, the complete roster of data available for analysis is cited. This report is based upon the initial phase of analysis of selected portions of this body of data; further analysis is currently underway.

Results

The results here reported are based exclusively on New York State Readiness Test data obtained from children just beginning first grade. This test is made up of six subtests: Word Meaning, Listening, Matching, Alphabet, Numbers and Copying. The Word Meaning Test (16 items) asks the child to mark one of three pictures in each row named by the examiner. All but one of the items involve nouns.

The Listening Test (16 items) involves listening to an orally presented detailed statement of a sequence of events or a description of a more or less familiar object. Usually all that is required is close attention to and retention of the details of the content, and then relating them to the appropriate pictorial representation of this same content. In a number of instances, however, relatively unfamiliar objects or events are alluded to so that under these circumstances, the

item primarily calls for the availability of information rather than listening.

The Matching Test (14 items) requires the child to find the figure among three alternatives that is exactly the same as a comparison figure. Eight of the 14 items present words, or rather, from the children's viewpoint, sequences of letters, and the remaining items present abstract geometric figures. The Alphabet Test requires the child to mark the letter read aloud by the examiner.

The Numbers Test contains the largest number of items and is the most heterogeneous in content. It calls for simple forms of ordination and cardination, writing and recognizing numbers, knowing about money, and using arithmetic reasoning to solve simple problems. The Copying Test requires the child to copy a set of 14 figures. The first four items presented letters or numerals while the remainder involved geometric figures of varying complexity.

The mean proportion of items answered correctly for each subtest in each school is given in Table 2 along with relevant normative data. There is great consistency to the pattern of test results. In all three high scoring middle-class groups, the highest proportion of items answered correctly was achieved in the Word Meaning, Listening and Alphabet subtests. Among the low scoring, disadvantaged groups, the highest proportion by a substantial margin was in the Listening subtest, with Word Meaning usually ranking second. In all the groups, irrespective of whether they were high or low scoring, the lowest proportion, by a wide margin, was obtained on the Copying subtest.

Relative to the mean and quartiles of the norms set by the standardization group, the three high scoring school groups performed best on the Word Meaning and Listening subtests. Lagging somewhat, but still exceeding the means by a considerable margin, are their mean scores on the Numbers and Alphabet subtests. Their Matching subtest mean scores are just above the means of the standardization group. Mean scores on the Copying test trail badly.

The mean subtest scores for each of the five low scoring school groups fell just short of the mean of the normative group on the Listening subtest. The means of these same groups fell approximately halfway between the first and second quartiles of the norms on the Word Meaning subtest. On the Matching subtest, their mean scores were usually somewhat above the first quartile, whereas in the Alphabet subtest, their mean scores fell very close to those corresponding to the first quartile. In four out of the five cases, the mean score on the Number subtest was substantially below the first quartile. All but one of all nine school groups, including the four high scoring groups, scored below the first quartile on the Copying subtest, suggesting that for this particular subtest at least, the norms are in serious error -- or that the administration and scoring of this subtest in the New York City School System are uniformly different from those specified in the test manual.

Thus it would appear that all the school groups, irrespective of their overall performance, performed best on the first two subtests -- Word Meaning and Listening -- and poorest by far on the Copying subtest. Relative to the range of scores obtained on each of the subtests, it would appear that the smallest differences found between the high and low scoring groups were in the Listening and Matching subtests (ignoring the findings on the Copying subtest because of their skewness in relation to the published norms). The greatest difference between the high and low scoring groups was found on the Numbers subtest.

Analysis of school group differences by subtest:

Test 1. Word Meaning

From the proportions obtained on three practice items that preceded the actual test, it would appear that most of the children understood the relatively simple instructions associated with this subtest. Clear differences were obtained between

the high and low achieving groups on all of the first 11 of the 16 items on this subtest. There is a consistent pattern differentiating the performance of these school groups. More than 50% of the children from each of the three high scoring schools answered 10 of these first 11 items correctly, whereas only two of these 11 items were answered correctly by more than half of the children in each of the five low achieving school groups. On the remaining nine items, there were only two instances of any of the five school groups obtaining a proportion correct score exceeding .5. The uniformity of the item performance patterns, and the fact that the children from the low scoring school groups functioned well on two of the subtest's 16 items, would tend to suggest that most of the children understood the task* and were able to maintain sufficient interest to respond appropriately when they knew the correct answer. On the basis of the admittedly small sample of words included in the 16-item Word Meaning subtest, it would appear that there are very large differences between the two school groups in the number of words whose meaning they understood, and/or in their ability to recognize pictorial representations associated with words.

Test 2. Listening

The data indicate clear regularities in performance from school to school as a function of item difficulty. The difference between the high scoring and low scoring school groups tend to be consistent but much more modest than in the Word Meaning subtest; there is no dramatic gap between the high and low achievers on this subtest.

Test 3. Matching

Scores on this subtest were substantially affected by the position of the correct alternative. Performance was best on those items in which the correct

---

*There is, however, clear indication that some of the children in the low scoring groups simply did not know how to respond to one or all of the subtests. Tabulation of these instances has not yet been done.

alternative is presented first. Apparently, when the correct alternative was contiguous with the comparison figure, comparison was greatly facilitated. There were large differences in performance between the high and low achieving school groups, but substantial variation between schools within the disadvantaged group obscured some of the differences between the high and low groups. Although all the school groups were affected by the sequence of alternatives, in many of the low scoring school groups, there was little if any success when the alternative was not contiguous with the comparison figure. The sequence of the alternatives rather than the content of the comparison figure (whether it consisted of words or abstract designs) appeared to be the decisive determinant of success, suggesting that relatively few children had sufficient reading skills to affect their performance on the word items.

Test 4. Alphabet

There are large consistent differences in performance between the high and low scoring groups. Over 60% of the children in the high scoring school groups were usually able to identify each of the letters, whereas there was seldom more than one out of three children in the low scoring school groups who were able to identify the letters correctly. Although the items varied somewhat in their difficulty level, the major source of variance was the social class background distinction of the groups.

Test 5. Numbers

In addition to the problem of the heterogeneity of its content, this test contains peculiar format features. First, the items occupy areas of different size on the answer sheet, so that it may have been difficult for some children to find the appropriate item on the page when the examiner was reading the questions. Second, there are several items in which it is not exactly clear to the examiner how he should indicate his answer. Finally, some of the problems involve rather detailed

oral statements by the examiner, so that part of the test, by virtue of its content, should more appropriately appear in the Word Meaning or Listening subtests.

Despite these obfuscating features, rather consistent differences between the high scoring and low scoring school groups are apparent. Virtually all the children could point to the biggest apple among four of varying size (item 1). Most of the children could identify the watch whose hand was pointing toward three, but in this instance there were substantial differences between the middle and lower-class groups.

Three items called for the child to write a number -- a one, two and three-digit number. The one-digit number was considerably easier than the two-digit number, and the three-digit number was impossible for all but a handful of children to do. The preponderance of middle-class children were able to write the one-digit number, whereas only about one of three disadvantaged children had this skill. A greater proportion of high achieving children were able to write two-digit numbers than low achieving children were able to write single-digit numbers; two-digit numbers were only rarely made by children from the low achieving group. The very limited ability of these children to cope with three-digit numbers was further illustrated by their widespread failure to identify correctly the largest of four three-digit numbers presented to them. Only about three in ten, and a little more than one in ten from the high and low scoring groups, respectively, chose the correct answer on this item.

Similar differences between the high and low scoring groups were found in their ability to recognize one and two-digit numerical figures. In fact, slightly more middle-class children could actually write a one or two-digit number than could disadvantaged children recognize a one or two-digit number. For all groups, naturally, the task of recognizing a number was easier than writing it.

Three items presented a counting task, each under somewhat irregular conditions.

The easiest item called for the child to find a house with seven windows. The preponderance of children were able to perform this feat, whereas only a comparatively small number could identify the rectangle which had 12 dots dispersed unevenly throughout its interior. In both instances, the children from the high scoring school group performed somewhat better than those from the low scoring group. This discrepancy increased markedly in their performance on the third item, which called for the child to mark the "seventh bird from the nest" from among a row of birds adjacent to a nest. Apparently the wording of this item almost completely disabled the disadvantaged groups from functioning on this item, whereas more than half the middle-class children were able to pass it.

Four items involved money in some fashion. One asked the child to identify the coin that will buy the most (the correct answer, a picture of a quarter, was also the largest coin); another, to indicate which item (a car or various two or three-wheeled vehicles) costs the most money, while the remaining two items asked for the number of pennies in a dime and in a quarter. The first two items were much easier, especially the one about the coin with the greatest purchasing power. In both, middle-class children performed somewhat better than the disadvantaged groups. This gap widened in the case of the other two items. In these latter instances, however, fewer than half of the middle-class children were able to provide the right answers.

Several problems requiring simple arithmetical reasoning are included in this test. The high scoring groups rather consistently performed better than the low scoring groups, but only in three instances did more than half the children from any of the eight school groups succeed in passing any one of the six problems. Problems calling for simple addition were somewhat easier than those involving subtraction; those requiring multiplication or division were most difficult.

One problem calling for the identification of the glass that was half full was answered correctly by approximately half the high achieving school groups in

contrast with but a fourth of the low achieving groups. A similar problem requiring understanding of a fraction -- one fourth -- was only seldom answered correctly by any of the groups.

Two items that ask the child to find the number that is more than one designated number and fewer than another designated number, were seldom answered correctly. As stated on the test, the concept of an interval within which a given number may fall was unfamiliar to these children.

Test 6. Copying

Only the first four items of this subtest (the ones presenting letters or numerals) were handled effectively, more so by the high achieving groups. All eight of the school groups performed well below the median of the norms. Either the conditions of administration were deviant -- or more severe criteria of scoring were used, or there is a pervasive deficiency in the ability to perceive and reproduce geometric figures among New York City children.

Discussion

The results of this micro-analysis of Readiness Test performance indicate that there is a rather widespread difference in ability level between middle-class and disadvantaged children, even at the time when they are about to begin first grade. Unless the differences that were found to pervade virtually every item of every subtest can be completely ascribed to a basic difference in test-taking skills, the data of this study indicate that differences between disadvantaged and privileged children are not restricted to one or two dimensions of the cognitive domain, but rather extend to every area of intellectual functioning the test constructor attempts to assess. Because of the pervasive quality of the differences found among these groups of children, it is reasonable to assume that there is a general trait relating to test-taking effectiveness, whether it be attentiveness or perseverance or achievement motivation that distinguishes the two groups. Nevertheless, the manner in which the curves depicting item-by item performance consistently remain parallel to each other suggest that the specific content of each item, too,

influenced test performance.

Since the sampling of the cognitive domain which determined the content of the subtests comprising the Readiness Test appears to be arbitrarily selective and rather uneven, it is impossible to conduct a definitive comparative analysis of the cognitive functioning of six-year-old disadvantaged and middle-class children on the basis of this test alone. Nevertheless, the comparative data suggest some important points. First of all, it should be noted that the subtest with perhaps the heaviest verbal loading -- the Listening Test -- was the one on which the disadvantaged groups performed best (as indicated by the admittedly tenuous norms of the test itself). This subtest required the child to attend to and remember a rather involved set of verbal statements. His relative degree of success in this area would tend to suggest that the emphasis upon the so-called verbal deficit of the disadvantaged child may be misdirected. While the data from the Word Meaning test do indicate that the two groups of children differed greatly in the number of words they knew, it is instructive to observe that the disadvantaged child deals relatively effectively with situations in which he must listen to a flow of conversation by his teacher which is made up of relatively simple, functional elements of language.

Perhaps the most concrete and clearcut generalization that may be drawn from these test findings is that whereas most middle-class children can recognize the letters of the alphabet upon entering first grade, this ability is not present among most disadvantaged children. Unlike many of the other variables the Readiness Test attempted to measure, the nature of the discrepancy in performance on the Alphabet subtest between disadvantaged and middle-class children is so simple to delineate and to measure with precision that it seems to call for a compensatory educational program that provides training in this particular realm. However, the differences found in other parts of the Readiness Test -- in the Word Meaning subtest, in the ability to copy figures, and in the test of numerical knowledge

(which indicated that differences existed in their ability to recognize as well as to write numbers, in using numbers in relation to the value of money and objects one could buy with money, and in the ability to use whatever knowledge of numbers they had to work out simple arithmetic problems), all suggest that compensatory programs that focus upon one or two concrete, narrowly circumscribed areas of intellectual functioning have the potential for erasing a deficit in only a fragment of the spheres of the child's intellectual functioning. Unless it can be demonstrated that these particular areas have a highly facilitating and central influence upon other areas of deficit as well as on the learning that is to take place in the school setting, it may be predicted that such compensatory programs will fall short of expectations. In light of the pervasive quality of the deficits of disadvantaged children suggested by the results of the New York State Readiness Tests, it is more useful to identify the nature and source of the integrative and organizational attributes which have thus far impaired their development of intellectual functioning, and to plan a school program that is geared to promote growth in this more basic, integrative level of functioning.

Work to be Done

The first-level analysis here presented for the New York State Readiness Tests will be applied to the test data available from other age level children. Further, the analysis of all test data will be extended to include:

1. An analysis of variability as well as central tendency to determine the distribution of children at various levels of functioning for the two basic comparison groups.

2. Differentiation of those children who performed poorly from those who did not know how to deal with the test at all.

3. Analysis of patterns of intra-individual variation in those instances where scores on different tests are available for the same child (grade 1: subtests

of Readiness Test; grade 3: reading and arithmetic tests).

4. Study of sex differences in performance at all age levels and the two social class groups.

5. Qualitative analysis of performance through the study of patterns of wrong responses.

6. Study of the effects of kindergarten attendance by disadvantaged children on first-grade Readiness Test scores.

7. Study of the effects of ethnic background on test performance along the lines recently initiated by Lesser.

Table 1

Roster of Available Test Data

School	Mean Reading Level for Selected Grades		Number of Cases upon which the Analysis of Test Data will be Based Metropolitan Achievement Tests				
	Grade 2	Grade 5	New York State Readiness Tests (Grade 1)	Primary 1 Reading Test Form B (Grade 2)	Upper Primary Reading Test Form B (Grade 3)	Upper Primary Arithmetic Test Form B (Grade 3)	Elementary Arithmetic Test Form B (Grade 3)
P.S. 324 Bronx	4.3	9.7	234	228	185	24	185
P.S. 472 Brooklyn	3.9	7.7	150	159	168	49	115
P.S. 599 Queens	3.9	8.2	70	28	76	25	74
P.S. 640 Manhattan	3.6	8.6	---	118	136	39	93
P.S. 332 Bronx	2.2	4.9	108	143	111	86	58
P.S. 484 Brooklyn	1.9	3.8	112	110	93	90	---
P.S. 550 Queens	2.0	4.1	106	111	79	56	24
P.S. 688 Manhattan	2.0	4.4	137	133	139	110	34
P.S. 692 Manhattan	2.0	4.0	89	134	120	99	28
	Normal 2.8	5.8	TOTAL: 1006	1164	1107	578	611

Table 2

Proportion of Items Answered Correctly on Each Subtest of New York State Readiness Tests

	N	Word Meaning	Listening	Matching	Alphabet	Numbers	Copying
			(High Scoring)				
P.S. 324	234	.70	.66	.56	.73	.57	.26
P.S. 472	150	.63	.64	.57	.64	.51	.32
P.S. 599	70	.65	.61	.55	.58	.47	.26
			(Low Scoring)				
P.S. 332	108	.45	.54	.45	.35	.35	.20
P.S. 484	112	.46	.53	.34	.30	.29	.15
P.S. 550	106	.40	.53	.29	.29	.29	.19
P.S. 688	137	.39	.56	.38	.30	.29	.18
P.S. 692	89	.44	.55	.29	.33	.29	.10
Normative Data:	Mean:	.54	.56	.54	.59	.46	.49
	Q_3:	.69	.69	.79	.88	.58	.71
	Q_1:	.38	.44	.29	.31	.35	.29